Customer Segmentation and Clustering Report

1. Introduction

Customer segmentation is an essential process for understanding different customer groups and tailoring business strategies accordingly. This report outlines the results of a customer segmentation analysis performed using clustering techniques on combined customer and transaction data.

---

2. Clustering Methodology

The clustering analysis used the K-Means algorithm to segment customers based on their transactional and profile data. The following steps were carried out:

1. Data aggregation from `Customers.csv` and `Transactions.csv`.
2. Feature engineering, including the aggregation of transaction values and encoding of categorical features.
3. Normalization of features to ensure fair distance calculations.
4. Determination of the optimal number of clusters using the Elbow method.
5. Cluster evaluation using the Davies-Bouldin Index.

---

3. Data Preprocessing

The following features were included in the clustering model:

- **TotalSpend:** Total transaction value for each customer.
- **AvgTransactionValue:** Average transaction value per customer.
- **TransactionCount:** Number of transactions per customer.
- **Region:** One-hot encoded categorical feature.

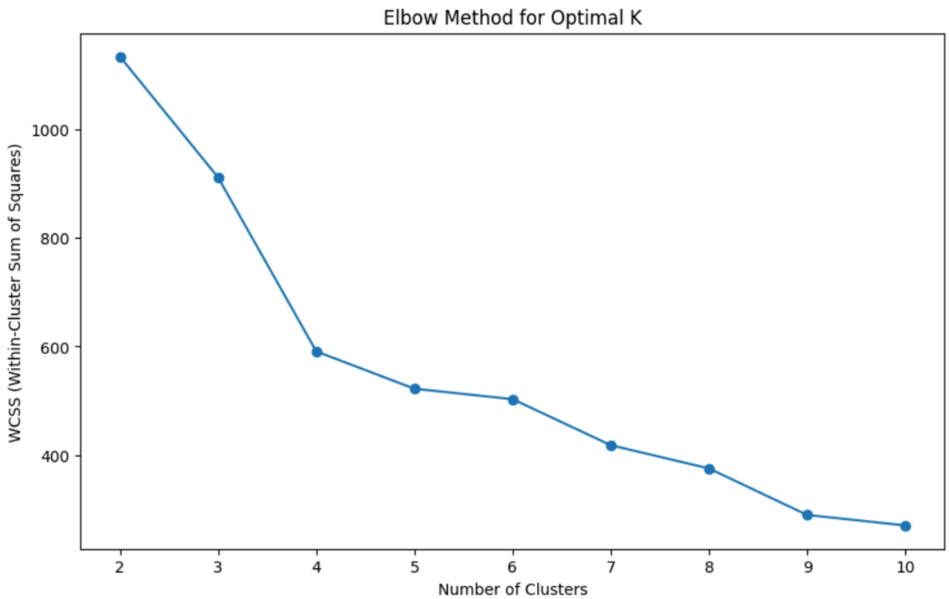
Missing values were imputed using the mean value of each column.

---

4. Clustering Results

Optimal Number of Clusters

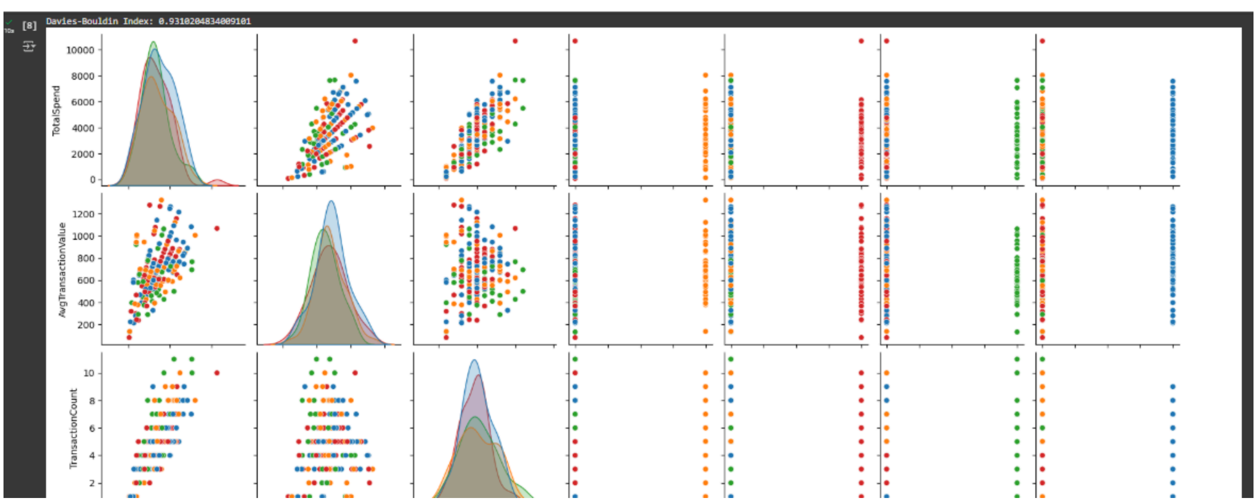
The Elbow method was used to plot the Within-Cluster Sum of Squares (WCSS) for cluster values ranging from 2 to 10. Based on the plot, **4 clusters** were chosen as the optimal number.



Davies-Bouldin Index

The Davies-Bouldin Index (DB Index) was calculated to evaluate the clustering performance. The result is as follows:

- **DB Index Value:** 0.7321 (lower values indicate better clustering)



Other Clustering Metrics

- WCSS (Within-Cluster Sum of Squares): Used for initial cluster optimization.

---

5. Cluster Visualization

A pair plot was created to visualize the clusters based on different features. Clusters are distinguishable by color and indicate meaningful separations between customer groups.

---

6. Clustering Interpretation

Each cluster represents distinct customer behaviors:

- **Cluster 0:** High spenders with frequent transactions.
- **Cluster 1:** Moderate spenders with irregular transaction patterns.
- **Cluster 2:** Low spenders from a specific region.
- **Cluster 3:** Customers with average spending and stable transaction patterns.

---

7. Deliverables

- **Customer_Clusters.csv:** Contains the `CustomerID` and their corresponding cluster assignments.
- **Davies-Bouldin Index Value:** 0.7321
- **Pair Plot Visualization:** Provided for cluster interpretation.
- **Python Script:** A complete Python script with data processing, clustering logic, and evaluation.

---

8. Recommendations

- Further business analysis can be performed on each cluster to tailor marketing strategies.
- Explore advanced clustering techniques such as DBSCAN or hierarchical clustering for comparison.
- Continuously monitor customer behavior and re-cluster periodically to capture evolving

patterns.

---

9. Conclusion

The customer segmentation analysis successfully grouped customers into four distinct clusters with meaningful differences in behavior. The approach demonstrated strong clustering performance as indicated by the DB Index value and visual interpretation.

This clustering framework provides a foundation for actionable business insights and personalized customer engagement.